

Recibido / Received: 11/11/2019

Aceptado / Accepted: 16/03/2020

Para enlazar con este artículo / To link to this article:

<http://dx.doi.org/10.6035/MonTI.2020.ne6.7>

Para citar este artículo / To cite this article:

León Araúz, Pilar & Melania Cabezas García. (2020) "Term and translation variation of multi-word terms." In: Mogorrón Huerta, Pedro (ed.) 2020. *Análisis multidisciplinar del fenómeno de la variación fraseológica en traducción e interpretación / Multidisciplinary Analysis of the Phenomenon of Phraseological Variation in Translation and Interpreting*. *MonTI Special Issue 6*, pp. 210-247.

TERM AND TRANSLATION VARIATION OF MULTIWORD TERMS

LA VARIACIÓN TERMINOLÓGICA EN LAS TRADUCCIONES DE LOS TÉRMINOS COMPUESTOS

PILAR LEÓN ARAÚZ

pleon@ugr.es

Universidad de Granada

MELANIA CABEZAS GARCÍA

melaniacabezas@ugr.es

Universidad de Granada

Abstract

Phraseology is central to specialized language. In scientific and technical communication, multiword terms (MWTs) (e.g. *volatile organic compound*) are the most frequent type of phraseological units. Rendering them into another language is not an easy task due to their cognitive complexity, the proliferation of different forms, and their unsystematic representation in terminographic resources. This often results in a broad spectrum of translations for MWTs, leading to higher term variation as a result of their composition by two or more constituents. In this study we carried out a quantitative and qualitative analysis of English term variants of MWTs from the environmental domain and their translations into Spanish. The focus was on translation variation and its occurrence in different linguistic resources.



Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

Keywords: Multiword term; Term variation; Translation; Corpus linguistics; Terminographic resource.

Resumen

La fraseología es una parte fundamental del lenguaje especializado. En la comunicación científico-técnica, los términos compuestos (p. ej. *compuesto orgánico volátil*) son las unidades fraseológicas más habituales. Trasladarlos a otra lengua presenta dificultades debido a su complejidad cognitiva, la proliferación de diferentes formas y su representación inconsistente en los recursos terminográficos. Ello suele derivar en un amplio abanico de traducciones para estos términos compuestos, que de por sí son proclives a la variación terminológica debido a su formación por dos o más elementos. En este trabajo realizamos un análisis cuantitativo y cualitativo de las variantes terminológicas de términos compuestos en inglés pertenecientes al ámbito medioambiental y de sus traducciones al español. En concreto, dedicamos especial atención a la variación en la traducción y su cobertura en diferentes recursos lingüísticos.

Palabras clave: Término compuesto; Variación terminológica; Traducción; Lingüística de corpus; Recurso terminográfico.

1. Introduction

Phraseology has been extensively studied in both general language (Firth 1957; Sinclair 1991; Mel'čuk et al. 1995; Corpas Pastor 1996, 2000; Mogorrón-Huerta 2010; *inter alia*) and specialized discourse (Picht 1991; Meyer & Mackintosh 1996; *inter alia*). The guiding principle of phraseology is that language is not only composed of individual words, but also of larger structures that function as a whole (Sinclair 1991). In their broadest sense, phraseological units are lexical units formed by two or more words that frequently co-occur and exhibit a variable degree of lexicalization, syntactic and semantic stability, and a possible idiomatic nature (Gläser 1988; Corpas Pastor 2000). There are different types of phraseological units (e.g. idioms, collocations, proverbs, etc.), whose consideration as such depends on the scope of the approach and how they comply with these features.

Specialized phraseological units have been defined as phraseological units conveying a specialized meaning which are frequently used in a scientific or technical domain, contain at least one term, and exhibit a certain degree of

lexicalization (Bevilacqua 2004: 28), e.g. *to generate power*. Specialized phraseological units are fundamental since they give semantic precision to terms, play a central role in conceptual systems, and allow expression according to the conventions of the different specialized domains. However, they have been less studied than their homologues in general language.

Multiword terms (MWTs) are the main type of phraseological unit in specialized discourse (Meyer & Mackintosh 1996; Ramisch 2015). In MWTs, two or more lexemes converge to form a new unit of meaning (e.g. *shrouded wind turbine*). These lexemes can join and form a graphic compound (*sacacorchos*) or keep the space in between (*pez espada*). Although the designation of *compound* is often restricted to the first type (Corpas Pastor 1996), we consider both types as compounds, even though our study focuses on separated or syntagmatic compounds.

Some authors refuse to consider MWTs as phraseological units (Zuluaga 1975; García-Page 2008). However, we agree with those including them in the group (Benson et al. 1986; Pawley 2001; Ramisch 2015; *inter alia*), because they share the defining features of phraseological units: the formation by two¹ or more elements, the frequent co-occurrence, the functioning as a whole, and a certain degree of lexicalization. They differ from idioms, the undeniable example of phraseological unit according to narrow approaches (García-Page 2008), because they are less lexicalized and idiomatic (thus, more transparent), and they convey concepts.

Precisely because these combinations are less lexicalized than other phraseological units, they are especially inclined to term variation or the coexistence of different denominations. For example, the Spanish idiom *a tenor de lo establecido* is a very lexicalized structure that does not seem to admit term variation. On the contrary, MWTs such as *contaminación por ozono* show different term variants, such as *contaminación por ozono troposférico* or *contaminación fotoquímica*. This happens because MWTs allow lexical expansion by combining concepts in different ways that are feasible in a particular

1. One of the defining features of phraseological units is their formation by at least two elements. However, these can often converge in a graphic compound (e.g. *rompeolas*). This does not mean that they lose their phraseological nature (in fact, there are studies, such as Oltra Ripoll's [2018], which include monolexical units in their phraseological analysis). Nevertheless, these are not the focus of our study.

domain (Picht 1991). For this reason, they can be more prone to variation than general phraseological units or other fixed, specialized combinations.

Furthermore, knowledge transfer and market expansion highlight the need for specialized translation, which implies dealing with a vast quantity of MWTs. Rendering them into other languages is not exactly easy, in part due to their characteristics, which make them cognitively and structurally complex, as well as their unsystematic treatment in terminological resources. This results in a wide variety of translation solutions, some more adequate than others in each context, which evidence a high degree of term variation when translating MWTs.

We carried out a quantitative and qualitative analysis of the Spanish translations of a set of English MWTs and their variants, with a focus on translation variation. Our goals were: (1) to define a typology of term and translation variants of MWTs; and (2) to investigate term variation of Spanish MWTs in different contexts (translation, bilingual or multilingual lexicography, and original production in Spanish). Our results showed that MWTs exhibit an enormous degree of term variation of different characteristics, which is particularly present in translation scenarios (i.e. parallel corpora).

The rest of this article is organized as follows. Section 2 explains term variation and its presence in specialized discourse. Section 3 focuses on variation in translation contexts. Section 4 presents the materials and methods of the study, and Section 5 describes the results obtained. Particularly, in Section 5.1 we present a typology of term variation found in translation, and in Section 5.2 translation variation is compared across different resources. Finally, Section 6 summarizes the conclusions of this study and future research lines.

2. Term variation in terminology

Variation is an essential feature of all languages, even in specialized discourse. It can be conceptual, when it affects meaning, or denominative, as explored in this study, when different designations are used to name the same concept (e.g. *wildfire* and *rural fire*). Not only are monolexical units variable, but also phraseological units, in particular MWTs.

However, in specialized discourse, both phraseology and variation were obscured for a long time, namely because the General Theory of Terminology

(TGT, Wüster 1968) downplayed problematic aspects such as context, phraseology, and variation. In fact, whilst aware of the existence of these phenomena in specialized discourse, phraseology and variation were perceived as obstacles for effective expert communication. The General Theory of Terminology took thus a prescriptive approach in which a term was said to allude to only one concept, and a concept was named by only one term. The richness resulting from variation was thus artificially ignored for the sake of precision, even though variation often emerges for the same reason, since new ways of conveying meaning are constantly sought for.

Variation did not become a focus until the advent of the new theories of terminology, which formulated communicative and cognitive approaches, and acknowledged the variable nature of both terms and concepts (Cabré 1993; Temmerman 2000; Freixa 2006; León Araúz 2017). Although to a lesser extent compared to general language (Freixa 2006; Sanz Vicente 2011), specialized discourse exhibits a considerable degree of variation, which has been discovered by means of corpora (Fernández-Silva 2018).

MWTs are largely used to illustrate specialized term variation. These terms are formed by two or more elements: a head and one or several modifiers. The head often indicates the category to which the concept belongs, whereas the modifier often indicates the criterion for subdivision of the category (Bowker 1998: 487). Depending on the nature of the head (e.g. object, property, process, etc.), the modifier can specify different types of features (e.g. purpose, location, method, material, etc.) (*ibid*: 488). Although some MWTs are fixed, it is evident that most of them are very often prone to variation (e.g. *air/atmospheric/air-borne/airborne pollutant*), as has been explained in different studies (Bowker 1998; Fernández-Silva & Kerremans 2011; Daille 2017; Giacomini 2018; Gledhill & Pecman 2018; Cabezas García & Chambó in press; *inter alia*).

Initially, the use of one variant or the other might seem arbitrary. However, as shown by Rogers (1997: 219), there are certain systematic patterns of variation that need to be explained. In the same line, Bowker & Hawkins (2006) affirm that variation cannot be attributed to the carelessness of subject field experts, but rather to their desire for precision and the carefulness invested in their choice of expression. On the one hand, variation sometimes happens with a specific purpose (Bowker 1998; Fernández-Silva et al. 2009; Kerremans

2017; Freixa & Fernández-Silva 2017; Gledhill & Pecman 2018), and, on the other, it can also reveal the novelty of concepts (i.e. neologisms) (Cabré 1993; Picton 2011).

As stated by Candel Mora & Carrió Pastor (2012), discovering the causes or types of variation is important for both theoretical and practical reasons. From a theoretical perspective, it reflects the mental processes involved in the selection of one term over another. On a practical level, this information is helpful for terminologists or translators in production tasks, since they need to know when to use one variant over another and the reasons why it is the best choice in a particular context.

Traditionally, the reasons for variation have been user-based (resulting in temporal, geographic, or social variation) or usage-based (i.e. field, tenor, and channel) (Gregory & Carroll 1978). Nevertheless, additional reasons can be involved in the variation of term denominations. As Freixa (2006) states, causes for term variation can be (1) dialectal; (2) functional; (3) discursive; (4) interlinguistic; and (5) cognitive. Several of these causes can also cooccur.

Dialectal reasons are based on the geographical, chronological, or social origin of speakers. Functional reasons are related to field, tenor, and channel. Discursive reasons can lead to stylistic and rhetoric changes. For instance, Collet (2003) argues that MWTs contribute to text cohesion by molding their shape in different contexts. Freixa & Fernández-Silva (2017), and Fernández-Silva (2018) state that intratextual term variation, especially in the case of MWTs, facilitates cohesion thanks to its repetitive nature. Along these lines, Fernández-Silva (2016) points out that MWT variants that emphasize different perspectives of the same concept contribute to knowledge construction. Accordingly, Gledhill & Pecman (2018) argue that English N+of+N MWT variants (e.g. *release of plumes*) usually convey knowledge rich contexts, while N+N MWTs (e.g. *plume release*) introduce new information.

Reasons for term variation can also be interlinguistic, when two or more languages are in contact and influence each other; and cognitive, which are based on the different conceptualizations of reality or user motivations. Cognitive variants are thus the natural reflection of multidimensionality, that is, the nature of those concepts that can be organized according to different facets or dimensions (Bowker 1998). As Meyer & Mackintosh (1996) argue, MWTs are the ideal scenario for multidimensionality since different facets

can be shown in the modifiers (e.g. *photochemical smog* [cause], *summer smog* [time], *ozone smog* [agent]). Term and concept variation, far from being unrelated, are thus the consequence of the convergent influence of multidimensionality, context and dynamism in specialized domains (León Araúz 2017).

Apart from the causes of variation, it is also important to reflect on its consequences. According to Fernández-Silva et al. (2009), term variation can have no cognitive consequences when there is only a change in the form but not in the meaning (e.g. *marine product* and *sea product*). Alternatively, term variation can have cognitive effects, when there is a shift in perception along with the change in form (e.g. *sea product* and *fishing product*). In this line, some classifications based on the semantic distance of term variants have emerged, such as the ones presented in Aguado de Cea & Montiel Ponsoda (2012) and Fernández-Silva (2018). Both studies distinguish three main groups: variants with (1) minimum, (2) medium, and (3) maximum semantic distance.

In the first group, the semantic content does not change, that is, terms are conceptually equivalent. For Aguado de Cea & Montiel Ponsoda (2012), this set includes synonyms, such as graphical and orthographical variants (*localization*, *localisation*), inflectional variants (*cat*, *cats*), and morphosyntactic variants (*nitrogen fixation*, *fixation of nitrogen*). Fernández-Silva (2018) adds morphological variants (*ozone-depletion potential*, *ozone-depleting potential*) and specifies that, in MWTs, synonymy can affect just one of the constituents (*organic matter pollution*, *organic matter contamination*).

For Aguado de Cea & Montiel Ponsoda (2012), variants with a medium semantic distance are partial synonyms or terminological units that highlight different aspects of the same concept, such as stylistic or connotative variants (*man*, *bloke*), diachronic variants (*tuberculosis*, *phthisis*), dialectal variants (*gasoline*, *petrol*), pragmatic or register variants (*headache*, *cephalalgia*), and explanatory variants (*immigration law*, *law for regulating and controlling immigration*). Fernández-Silva (2018) studies medium semantic distance in MWTs where the conceptual change is reflected in the modifiers. They can be subject to reductions (*motor vehicle emission*, *vehicle emission*), additions or deletions of non-defining characteristics (*non-point pollution*, *non-point source pollution*), or the use of a different defining feature (*summer smog*,

Los Angeles smog). Again, the implication of some of these characteristics is directly linked to multidimensionality and cognitive variants.

Finally, variants that entail a maximum semantic distance are terminological units that highlight different features of the same concept which belong to different conceptualizations, or variants that refer to two conceptually related concepts (Aguado de Cea & Montiel Ponsoda 2012; Fernández-Silva 2018). In the specific case of MWTs, they involve changes in the MWT head or even in both constituents. These variants are the most cognitively complex since the semantic category of the concept is altered. This change can be just an emphasis of a specific perspective of the conceptualization (*exhaust emission, exhaust pollution*). Alternatively, a hypernymic category (*doubly fed induction generator, DFIG machine*) or, on the contrary, a hyponymic category (*high level ozone, stratospheric ozone*) can be employed. It is also possible to replace both the head and modifier with totally different categories (*oil pollution, discharged hydrocarbon*), which is the most semantically distant type of variation.

Evidently, term variation can acquire a wide range of forms, as suggested in the examples above. The following list shows the classification proposed in Faber & León Araúz (2016: 12-13)², which encompasses different proposals found in the literature and offers a thorough picture of term variation types, specifying whether semantics or communicative situations are affected:

- (A) Orthographic variants that are not influenced by geographic origin and do not alter semantics or the communicative situation, e.g. *groundwater, ground water*.
- (B) Diatopic variants:
 - (i) Orthographic variants that do not modify semantics, e.g. *fecal, faecal*.
 - (ii) Dialectal variants, which can alter semantics if cultural factors are involved, e.g. *gasoline, petrol*.
 - (iii) Culture-specific variants, which affect semantics and the communicative situation, e.g. *dry lake, sabkha*.

2. Some of the examples have been changed with variants found in this study.

- (iv) Calques, which can modify semantics and the communicative situation, e.g. *environmentally hazardous substance* > *sustancia ambientalmente peligrosa*, *sustancia peligrosa para el medio ambiente*.
- (v) Borrowings, which can alter semantics and the communicative situation and can be adapted or not, e.g. *smog* > *smog*, *esmog*
- (C) Short form variants, which have an effect on the communicative situation:
 - (i) Abbreviation, e.g. *greenhouse gas*, *GHG*.
 - (ii) Acronym, e.g. *laser*, *Light Amplification by Stimulated Emission of Radiation*.
- (D) Diaphasic variants:
 - (i) Scientific variants, which influence the communicative situation:
 - Scientific names, e.g. *Dracaena draco*, *drago*.
 - Expert neutral variants, e.g. *Ocellaris clownfish*, *Amphiprion ocellaris*.
 - Jargon, e.g. *lap-appy*, *laparoscopic appendectomy*.
 - Formulas, e.g. *carbon dioxide*, *CO₂*.
 - Symbols, e.g. €, *euro*.
 - (ii) Informal variants, which affect the communicative situation and can also influence semantics:
 - Lay user variants, e.g. *Dragon tree*, *drago*.
 - Colloquial variants, e.g. *motor vehicle pollution*, *car pollution*.
 - Generic variants, e.g. *pollution*, *contamination*.
 - (iii) Domain-specific variants, which can modify semantics or the communicative situation if the domains have different term preferences, e.g. *mud* and *sludge* represent the same concept except that the former is used in Geology and the latter in Water Treatment.
- (E) Cognitive variants, which are usually MWTs:
 - (i) Dimensional variants, which alter semantics because they express different dimensions of the same concept, e.g. *esmog fotoquímico* [*photochemical smog*], *niebla tóxica estival* [*summer smog*].
 - (ii) Intentional variants, which have an effect on the semantics and/or the communicative situation, since they are used depending on the receiver's reaction sought, e.g. *climate change*, *climate emergency*.
- (F) Metonymic variants, which affect semantics by alluding to a part or material of the concept, e.g. *accidental water pollution*, *accidental marine pollution*.

- (G) Diachronic variants, e.g. *anhídrido carbónico* [*carbonic anhydride*], *dióxido de carbono* [*carbon dioxide*].
- (H) Non-recommended variants, e.g. since *mental retardation* now has negative connotations, it has been substituted by *intellectual disability*.
- (I) Morphosyntactic variants, which do not usually influence semantics but depend on the communicative situation, as well as on term preferences and collocations, e.g. *contaminación acústica* [*acoustic pollution*], *contaminación de ruido* [*noise pollution*].

The nature and scope of variants are very diverse and can have different consequences in communication. Nevertheless, terms can activate more than one variant type, which might make term choice more difficult. For example, *H2O* and/or *water* can be domain-based variants since the first one is more frequently used in Chemistry and Water Treatment domains than in Oceanography, for example. However, their use also depends on the communicative situation (i.e. formal or informal). On the contrary, the same type of variant can be expressed by more than one term. Diaphasic variants, in particular, form a continuum from more formal to informal (e.g. *thermal low pressure system*, *thermal low*, *thermal trough*, and *heat low*) (Faber & León Araúz 2016).

3. Term variation and translation

Term variation has also been explored in interlinguistic contexts. The notion of ‘equivalence’ becomes thus crucial since terminologists and translators often follow different criteria. While terminologists usually understand equivalence at the term level, because their goal is the inclusion of terms in terminographic resources, translators look for correspondence at the sentence or text level. That is why they search for the functional equivalence of their translations (Reiss & Vermeer 1984; Nord 1997) instead of a direct term correspondence, which would be the case for terminologists. In this line, Gerzymisch-Arbogast (2008) highlights the fact that translation is a text-based rather than a corpus-based activity, and Kerremans & Temmerman (2016) state that translators are not always guided by the principle (often adopted for structuring multilingual terminology databases) that a term in the source language should be rendered as a direct (literal) equivalent in the

target text. Therefore, terminological equivalence does not always correspond to translation equivalence (Kerremans & Temmerman 2016: 59).

Equivalence is thus a broader concept for translators than for terminologists and allows for a wide spectrum of translation mechanisms. Consequently, equivalents at the sentence or text level can be those reproducing the same function or effect than the source text (Reiss & Vermeer 1984; Nord 1997), rather than just those conveying the same concept. For this reason, the use of hypernyms or other variants reflecting different conceptualizations (e.g. as a result of multidimensionality) can be justified in a translation equivalence context.

Along these lines, understanding and finding equivalence for phraseological units can be problematic for translators. In the case of MWTs, comprehension problems may be related to the identification of the term as a single unit of understanding, its internal dependencies, or the semantics of its formants. In turn, production-related problems include the order in which its formants should be translated, the prepositions that should accompany them, or the distinction among different variants should they be found in any resource. The well-known notions of *translation problem*, *error*, *strategy*, and *technique* become thus central. A phraseological unit poses a *translation problem* when it “puts up resistance to being translated” (Oltra Ripoll 2018: 102). This can lead to *translation errors* or the inadequate resolution of a problem (Hurtado 2001: 279). Therefore, the analysis of term variation in translation, as the one carried out in this study, must consider the *strategies* or mechanisms employed to solve translation problems, and, specifically, the *techniques* or procedures to find the solution to a particular translation problem (e.g. omission, addition, modulation, paraphrase, etc.). As for the identification of translation equivalences, parallel corpora have traditionally been used (Déjean & Gaussier 2002; Daille & Morin 2005). These are collections of original texts and their aligned translations, which facilitate equivalent identification. However, they are scarce, particularly in specialized languages and in language pairs in which English is not involved, and the influence of the source language is obvious. These corpora are useful for error analysis or the identification of the different translation options, but complicate the study of idiomatic language uses. In consequence, comparable corpora are increasingly present, which consist in collections of texts of

the same type and domain, originally written in every language. These are more easily obtained and allow for the study of real language uses, although equivalent identification is evidently more complicated.

One of the studies devoted to term variation in translation is Fernández-Silva et al.'s (2009), who investigate, among other aspects, the role of the cultural system by means of term variants in French and Galician. Kerremans (2010, 2016) also studies term variation in specialized translation, with a focus on the reflection of the English source language variants in the target languages (Dutch and French). Along these lines, Fernández-Silva & Kerremans (2011) study cognitive term variants, and affirm that source language variants in Galician are reflected in the English target texts. Miyata & Kageura (2016) argue that translated texts (from Japanese into English) show a higher density of term variants as a result of the different translations possibilities. This finding was also confirmed in a previous study by Sanz Vicente (2011), which focused on the translation of English MWTs into Spanish and observed the higher coexistence of term variants in the target language. Accordingly, Jiménez-Crespo & Tercedor-Sánchez (2017) explore term variation in translated (English > Spanish) and non-translated texts (Spanish), paying particular attention to register, determinologization, explicitation, and term variation in translated documents. Consciously or unconsciously, these studies often use MWT examples to illustrate term variation in translations.

As some of these research studies suggest, MWT variation can influence their translations, where a higher degree of variation is usually found. Every language has different degrees of variation depending on their own linguistic characteristics. For instance, English is prone to graphical variation due to the use of hyphens, among other factors, whereas Spanish is more prone to morphosyntactic variation due to its compounding rules. For this reason, ascertaining term variants in the source language is relevant with a view to facilitating equivalent identification. In this sense, Rogers (1997) underlines that translators are frequently obliged to make decisions not only regarding synonymy within the source text and target text, but also regarding the cross-linguistic relations between these synonyms.

The representation of the different translation possibilities in terminographic resources becomes thus central in descriptive settings (Kerremans 2010), contrary to what has traditionally been done. These resources often

describe just a small portion of variants (if at all), which are represented in an unsystematic way. As for MWTs, those formed by more than two constituents are rarely included (Giacomini 2018). Users are often confronted with a lack of information on how term variation arises and which selection criteria to choose. Evidently, they need to know when to use each variant as well as its conceptual and communicative implications, since this will affect the receiver's interpretation of the message. Otherwise, translators can actually over-standardize, thus creating consistency in places where the use of variants was deliberate and well-reasoned (Bowker & Hawkins 2006: 80). Translators and terminologists must thus learn to recognize the patterns that lie beneath variation so that they do not inappropriately standardize terminology (*ibid*: 101). However, while refusing the wüsterian principle of univocity, we believe that the proliferation of term variants in target texts is often puzzling and overwhelming, partly due to an unsystematic treatment of MWTs, since translators do not always consider them as a single unit of understanding. Consequently, besides describing different types of variants, which is undoubtedly important, the added value of a linguistic resource lies in the enhancement of those data with additional information, such as semantic, pragmatic, and usage aspects (Faber & León Araúz 2016; Giacomini 2018), which improves a sound use of variation in texts.

4. Materials and methods

According to Kerremans & Temmerman (2016: 45), corpora allow us to study the textual and linguistic features of translations, taking into account different contextual parameters that have an impact on translation choices and, ultimately, on the translation product. For this reason, different corpora were used for MWT extraction and term variant identification.

Since one of our goals was to explore MWT variation patterns in translation situations, the OPUS2 English corpus (Tiedemann 2012) was first used to extract a set of MWTs, which would then be compared with their equivalents in the OPUS2 Spanish parallel corpus (Tiedemann 2012). The OPUS2 English corpus is an open source parallel corpus that can be accessed at Sketch Engine (<https://www.sketchengine.eu/>) (Kilgarriff et al. 2014) and encompasses 40 languages. It is organized in subcorpora, such as the European Central

Bank (ECB), the European Parliament Proceedings (EUROPARL), and the Translated UN documents (MultiUN). Aware of the scarcity of specialized parallel corpora, we selected this parallel corpus, which includes both general and specialized corpora, and decided to focus on specialized terms of general interest. This is the case of POLLUTION, a specialized concept that is present in everyday communication due to the increasing climate awareness.

Therefore, starting from the term *pollution*, a conceptual analysis was carried out in the corpus, which allowed us to identify pollution-related concepts, such as *emission*, *ozone*, *gas*, *pollutant*, *substance*, *contamination*, *smog*, *air*, etc. These terms were then used as MWT heads in CQL (Corpus Query Language) queries in Sketch Engine, which allow to search for specific morphosyntactic patterns, such as MWTs premodified by different elements (Table 1):

[tag="N.* JJ.* RB.* VFN.* VVG.*"]{1,}[lemma="pollution"][tag!="N.* JJ.*"]

Table 1: CQL query to extract MWTs whose head is *pollution*.

The CQL expression in Table 1 elicits premodified MWTs (the most frequent structure of these terms), such as *heavy metal pollution* or *indoor air pollution*. It searches for the lemma *pollution* ([lemma="pollution"]) (or *emission*, *ozone*, *gas*, *pollutant*, etc., in the following queries) preceded by nouns, adjectives, adverbs, past participles, or present participles ([tag="N.*|JJ.*|RB.*|VFN.*|VVG.*"]) appearing one or more times ({1,}). On the right of the head *pollution*, a restriction is included in order to exclude nouns or adjectives ([tag!="N.*|JJ.*"]), which could indicate that *pollution* is not the MWT head but takes part in a longer term.

This query was repeated for the other possible heads (*emission*, *ozone*, *gas*, etc.), and the 234 most frequent MWTs were selected. Several of these MWTs were term variants, and were thus grouped by concept. Finally, we obtained a set of 98 pollution-related concepts.

To identify the Spanish variants of the English MWTs, different resources were used. Firstly, with a view to investigating MWT variants in translation contexts, we used three parallel corpora, which allowed an easy identification of variants thanks to their alignment. Although, ideally, parallel corpora of specialized texts on the environment in English and Spanish would be the

best option, due to their scarcity we used parallel corpora encompassing both general and specialized discourse in a wide variety of domains.

These parallel corpora were the OPUS2 English-Spanish corpus (Tiedemann 2012), which was presented above in its monolingual version in English; the EurLex English-Spanish corpus (Vaisa et al. 2016), a multilingual corpus in all the official languages of the European Union that includes texts in the EUR-Lex database and is available in Sketch Engine; and Linguee (<https://www.linguee.es/>), an online corpus of aligned translations in different languages, such as English and Spanish, which includes general as well as specialized texts. Even though Linguee does not allow specific CQL queries and shows just a summary of the possible translations, it complemented the alignment mismatches that were often found in OPUS2 and EurLex. As a consequence, it allowed us to collect more different translations.

Secondly, in order to compare term variants found in translation contexts with those present in bilingual or multilingual lexicographic scenarios, Spanish equivalents of English MWTs were also looked up in two terminological databases: TERMIUM Plus and IATE. TERMIUM Plus (<https://www.btb.termiumplus.gc.ca/>) is a terminological database developed by the Government of Canada, which describes millions of concepts from specialized domains in English, French, and Spanish. Additionally, IATE (<https://iate.europa.eu/>) is the EU's terminology database and includes terms from a wide range of specialized domains in the official languages of the European Union. The entries consulted in these resources also allowed us to expand the collection of English source terms, since many of their entries contain synonyms. Therefore, a new set of terms was researched in the parallel corpora in order to expand the collection of Spanish translation variants. The final set of terms (a total of 277) ranged from two-word terms (e.g. *oil pollution*) to six-word terms (e.g. *aggregate anthropogenic carbon dioxide equivalent emissions*).

Finally, with a view to analyzing term variants in a context of original production in Spanish, we used a Spanish comparable corpus of specialized texts on the environment. The corpus was compiled by the LexiCon research group of the University of Granada while building EcoLexicon (<https://ecolexicon.ugr.es>, León Araúz, Reimerink & Faber 2019), a terminological knowledge base on the environment, and consists of approximately 10 million words. Since the size of the corpus cannot compete with the size of parallel ones,

this was compensated by the use of Google Scholar (<https://scholar.google.com/>) as a second comparable corpora.

Evidently, Google Scholar is not a comparable corpus strictly speaking. Moreover, it does not allow for flexible searches, such as lemmatized or CQL queries. Searches were time-consuming, since many variants needed to be looked up several times taking into account the inflectional rules of Spanish (i.e. number and gender). However, this was useful to obtain more results and measure the frequency of all variants found in the previous resources. We decided to only retain those terms that occurred a minimum of 10 times. At least in terms of frequency, this corpus should be more reliable due to its size. We agree with Bowker & Hawkins (2006), who state that the Web represents a huge and easily accessible body of linguistic data. Grefenstette (2002: 207) supports this claim by arguing that the size of the Web compensates for its “dirtiness”: “the correct form is always orders of magnitude more frequent than the erroneous form (...). The Web is dirty but the signal (correct forms and correct usage) is so strong noise can easily be ignored”. In this case, restricting the queries to Google Scholar ensures specialized language since Google Scholar is limited to research works.

It should be noted that the sequence of resources presented (i.e. parallel corpora, terminographic resources, and comparable corpora) was not random. We started with resources that provide direct access to interlinguistic variants (i.e. parallel corpora and terminographic databases), and then the last step was querying the comparable corpora, which required specific strategies to findequivalences the search process of which was more complex.

Our equivalence identification strategy in the comparable corpora involved the following queries. The terms found in parallel corpora (e.g. *contaminante atmosférico*, *contaminante aéreo* [*atmospheric pollutant*]) were literally searched for to confirm their presence in the comparable corpora. It should be noted that some of the variants obtained from parallel corpora were not queried in the comparable corpora, since they could bias the results. This was the case of hypernyms used as term variants. For instance, although *contaminación acústica* [*acoustic pollution*] and *ruido* [*noise*] are used as variants in the parallel corpora, searching for *ruido* [*noise*] in the comparable corpora would elicit additional meanings that would spoil the results. The

same was true of polysemic acronyms in Google Scholar (e.g. *GEI* [*GHG*]) and *ad hoc* variants that could not convey exactly the same meaning (e.g. *daños medioambientales* [*environmental damages*] as a term variant of *contaminación medioambiental* [*environmental pollution*]).

Additionally, the MWT heads and modifiers found in the parallel corpora were used with a span in between, as in the CQL query shown in Table 2. The span was set in 5 elements in order to allow for different possibilities without being too broad, since concepts in an MWT don't usually have a wider distance, as found in previous studies. However, in larger MWTs, such as those including participles or relative sentences, a higher span was used. Since CQL queries are not possible in Google Scholar, the * wildcard was employed to indicate the span. This allowed us to obtain new possibilities of extended MWTs, such as *contaminantes liberados a la atmósfera* or *contaminantes vertidos a la atmósfera* [*pollutants released into the air*].

[lemma="contaminante"][]{}0,5[lemma="aire atmósfera atmosférico aéreo"]

Table 2: CQL query for eliciting MWTs of the type *contaminante * aire/atmósfera/atmosférico/aéreo*.

Other queries in the specialized corpus of EcoLexicon included searching for the head and sorting the results by right so as to distinguish the different modifiers accompanying the head easily. The same strategy was applied to modifiers, sorting by left to discover the different possible heads. This type of queries is more time consuming, that is why they were only used in the EcoLexicon corpus, given the corpus size and the restriction possibilities, but not in Google Scholar. Figure 1 shows some of the modifiers that follow *sustancia* [*substance*] in the EcoLexicon Spanish corpus: *sustancia contaminante* [*polluting substance*], *sustancia explosiva* [*explosive substance*], *sustancia nociva* [*harmful substance*], *sustancia peligrosa* [*dangerous substance*], *sustancia que agota el ozono* [*ozone-depleting substance*], *sustancia que agota la capa de ozono* [*ozone-depleting substance*], its abbreviation *SAO* [*ODS*], *sustancia química* [*chemical substance*], and *sustancia tóxica* [*toxic substance*].

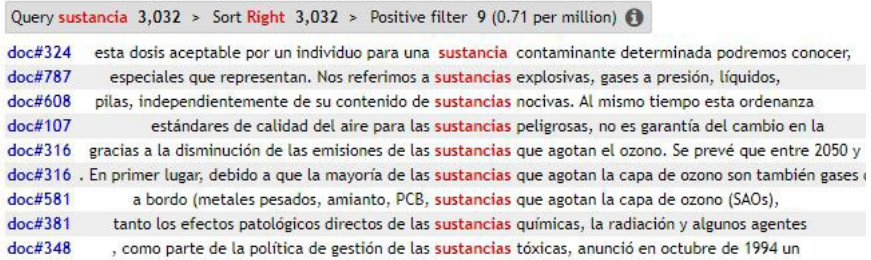


Figure 1: Sample of the modifiers of *sustancia* in the query *sustancia* + sort by right.

Evidently, concordance lines can also reveal knowledge-rich contexts that provide new term variants, as exemplified in the abbreviation SAO (*sustancia que agota la capa de ozono*) [ODS, *ozone-depleting substance*]. The following example also shows the usefulness of knowledge-rich contexts in variant identification, based on explicit knowledge patterns (*o*, the disjunctive conjunction): “Hemos de distinguir entre *contaminación natural* o *endógena* y *contaminación antrópica* o *exógena*”. As can be seen, *contaminación natural* [natural pollution] and *contaminación endógena* [endogenous pollution] are term variants and antonyms of *contaminación antrópica* [anthropogenic pollution] and *contaminación exógena* [exogenous pollution], another variant that had not been elicited in parallel corpora or terminological resources. Consequently, these strategies revealed new variants and complemented the queries in the parallel corpora and the terminological databases. In total, 1428 Spanish term variants were retrieved. Table 3 shows the English terms naming the concept OZONE POLLUTION and the Spanish variants found in the different resources.

OZONE POLLUTION	
English term variants	
ozone pollution tropospheric ozone pollution lower atmospheric ozone pollution surface ozone pollution	pollution by ozone photochemical pollution pollution caused by tropospheric ozone ground-level ozone pollution
Spanish translation variants	
OPUS2 Spanish	EurLex Spanish
contaminación por ozono contaminación fotoquímica ozono troposférico contaminante	contaminación por ozono contaminación por el ozono contaminación fotoquímica contaminación del ozono de la troposfera contaminación por ozono troposférico contaminación provocada por el ozono contaminación troposférica por ozono contaminación por ozono en la baja atmósfera
Linguee	Google Scholar
contaminación por ozono contaminación del ozono contaminación por ozono troposférico contaminación fotoquímica contaminación de ozono a nivel del suelo contaminación por ozono de la baja atmósfera contaminación del ozono a nivel del piso contaminación por ozono en la baja atmósfera	contaminación de ozono a nivel del suelo contaminación del ozono contaminación fotoquímica contaminación por el ozono contaminación por ozono contaminación por ozono en la baja atmósfera contaminación por ozono troposférico
EcoLexicon	IATE and Termium
contaminación por ozono contaminación fotoquímica	contaminación por ozono

Table 3: English variants for OZONE POLLUTION and Spanish variants found in the different resources.

5. Results and discussion

5.1. Characterizing translation variation

The analyzed concepts show a high degree of variation in Spanish, ranging from 2 term variants (e.g. *sectorial emission*) to more than 46 (e.g. *aircraft noise*). Few of them appear to be highly lexicalized, but among those who are, what seems to be a trend is the fact that having an acronym makes them more stable, such as *compuesto orgánico volátil (COV)* [*volatile organic compound, VOC*]. On the contrary, codifying a causal relation was found to make MWTs more prone to variation, since they usually present multiple periphrastic structures making the semantics of the concept explicit. One of these examples is *anthropogenic emissions*, which can be rendered as *emisiones antropogénicas*, but also as *emisiones procedentes de fuentes humanas*, *emisiones generadas por el hombre*, *emisiones causadas por el hombre*, *emisiones producidas por los humanos* or *emisiones provocadas por el hombre*, among other variants.

It is also worth noting that the more specific the concepts are, the more stable their designations were found to be, even if their hypernyms show many more variants. For instance, it is striking to see that while *contaminación atmosférica transfronteriza* [*transboundary air pollution*] can be rephrased as *contaminación transfronteriza del aire*, its hyponym *contaminación atmosférica transfronteriza a larga distancia* [*long-range transboundary air pollution*] does not follow the same pattern to be rephrased as *contaminación transfronteriza del aire a larga distancia* or *contaminación transfronteriza a larga distancia del aire* (not found in any of the resources). Thus, depending on the bracketing structure (i.e. internal dependencies in the MWT), the distance between the different elements of the MWT seems to present a limited span.

In what follows variation found in the different resources has been parametrized in a typology specifically conceived to characterize translation correspondences. Although many of the variants described in Faber & León Araúz (2016) occurred (e.g. orthographic variants, dialectal variants, calques, adapted and non-adapted borrowings, short form variants, etc.), most of them were morphosyntactic and cognitive, which calls for an expansion of this variation categories in translation scenarios. Partially based on traditional

classifications of translation strategies³ (Hurtado 2001; Newmark 1981, *inter alia*) and Daille's typology for French term variants (Daille 2003), our typology is divided in five main groups (omissions, structural shifts, transpositions, expansions and inaccuracies), each of them including other strategies, such as hyponymy, synonymy, paraphrases and modulation.

1. Omissions

a. Omission of articles

- e.g. *total de las emisiones agregadas de GEI*, *total de emisiones agregadas de GEI*
- e.g. *contaminación de las aguas subterráneas*, *contaminación de aguas subterráneas*

b. Omission of formants

i. Omission of modifiers

1. Omission implying activation of hypernym

- e.g. *contaminación atmosférica transfronteriza*, *contaminación atmosférica*
- e.g. *contaminación marina por petróleo*, *contaminación marina*, *contaminación por petróleo*

2. Omission not implying activation of hypernym

- e.g. *contaminación de fuentes difusas*, *contaminación difusa*
- e.g. *emisión de gases de efecto invernadero*, *emisión de gases invernadero*

ii. Omission of the head (sometimes implying transposition)

- e.g. *emisión de gases de efecto invernadero*, *gases de efecto invernadero*
- e.g. *sustancia tóxica*, *tóxico*

2. Structural shifts

a. Shift of prepositions

i. Prepositions expressing cause

3. The scope of traditional classifications is usually wider and not necessarily term-oriented. Furthermore, strategies are always described with regards to a source text, whereas in this study comparisons are established between different term variants within the same language.

- e.g. *contaminación por metales pesados, contaminación con metales pesados, contaminación de metales pesados*
- ii. Prepositions expressing location
 - e.g. *ozono en el nivel del suelo, ozono al nivel del suelo, ozono del nivel del suelo*
- iii. Prepositions expressing affected entity
 - e.g. *contaminación del agua, contaminación en el agua*
- b. Permutations
 - i. Permutation of modifiers
 - 1. Permutation with no meaning change
 - e.g. *emisiones antropogénicas de gases de efecto invernadero, emisiones de gases de efecto invernadero antropogénicas*
 - e.g. *contaminación marina accidental, contaminación accidental marina*
 - 2. Permutation with meaning change
 - e.g. *reducción certificada de emisiones, reducción de emisiones certificadas* (See Inaccuracies)
 - ii. Permutation of head and modifiers
 - e.g. *contaminación de fuentes no localizadas, fuentes de contaminación no localizadas*
 - e.g. *emisiones totales de GEI, total de emisiones de GEI*
 - *contaminante gaseoso, gas contaminante*
 - *contaminante atmosférico, aire contaminado*
- c. Shift of noun within modifier
 - i. Shift by synonym
 - e.g. *emisión de los tubos de escape de los automóviles, emisión de los tubos de escape de los coches*
 - ii. Shift by hypernym
 - e.g. *contaminación por metales pesados, contaminación por materiales pesados*
 - iii. Shift by metonym
 - e.g. *sustancia destructora de la capa de ozono, sustancia destructora del ozono*
 - iv. Shift by modulation (conceptual dimension)
 - e.g. *smog de verano* (time), *smog de Los Ángeles* (location)

- d. Shift of noun within head
- i. Shift by synonym
 - e.g. *aumento del nivel del mar*, *subida del nivel del mar*
 - ii. Shift by modulation
 1. Shift by near-synonym
 - e.g. *destrucción de la capa de ozono*, *empobrecimiento de la capa de ozono*
 2. Shift by metonym
 - e.g. *contaminación de las aguas subterráneas*, *contaminantes de las aguas subterráneas*
 3. Shift by conceptual dimension
 - e.g. *contaminación acústica*, *molestias acústicas*
- e. Shift of both modifiers and head (modulation)
- e.g. *contaminación causada por el ozono*, *polución urbana*
 - e.g. *smog fotoquímico*, *niebla tóxica estival*
 - e.g. *contaminación marítima por hidrocarburos*, *marea negra*
- f. Shift of modifying adjectives
- e.g. *contaminación no localizada*, *contaminación dispersa*
 - e.g. *emisiones antropógenas de gases de efecto invernadero*, *emisiones antropogénicas de gases de efecto invernadero*
 - *contaminación marina*, *contaminación marítima*
3. Transposition
- a. Transposition of adjective by periphrasis
 - e.g. *cambio climático antropógeno*, *cambio climático producido por el hombre*
 - e.g. *contaminación lumínica*, *contaminación producida por la iluminación*
 - e.g. *ozono troposférico*, *ozono en la baja atmósfera*
 - b. Transposition of adjective by “of + noun” (transposition)
 - e.g. *contaminante atmosférico*, *contaminante del aire*
 - e.g. *ozono superficial*, *ozono de superficie*
 - c. Transposition by inflection of one of the formants
 - e.g. *contaminación atmosférica transfronteriza a larga distancia*, *contaminación atmosférica transfronteriza a largas distancias*
 - e.g. *contaminación del agua*, *contaminación de las aguas*

4. Expansions

a. Making conceptual information explicit

i. Agent

- e.g. *contaminación de origen terrestre, contaminación por efluentes provenientes de tierra*
- e.g. *contaminación del ozono, contaminación provocada por el ozono*

ii. Patient

- e.g. *contaminación de origen terrestre, contaminación marina de origen terrestre*

b. Adding a second variant

- e.g. *ozono troposférico, ozono troposférico (al nivel del suelo)*
- e.g. *contaminación de fuente no puntual, contaminación de fuente no puntual (difusa)*
- e.g. *contaminación a larga distancia, contaminación a larga distancia o transfronteriza*

5. Inaccuracies (some of them can be considered translation errors [a-d] and some others as cognitive or intentional variants [e]).

a. Inaccuracies related to the semantics of one of the formants

- e.g. *contaminación atmosférica transfronteriza a larga distancia, contaminación atmosférica transfronteriza prolongada*
- e.g. *contaminación por nitratos, contaminación por nitrógeno*
- e.g. *contaminación por fuente no localizada, contaminación que no viene de fuente*

b. Inaccuracies related to the semantic relation between the formants

i. Confusion origin-patient

- e.g. *contaminación del terreno, contaminación de origen terrestre*

ii. Confusion patient-agent

- e.g. *contaminación en materias orgánicas, contaminación por materia orgánica*

c. Inaccuracies related to bracketing (i.e. internal dependence analysis)

- e.g. *contaminación del aire urbano, contaminación urbana del aire*

- e.g. *gases de efecto invernadero antropogénico, gases antropogénicos con efecto invernadero*
- e.g. *gases de efecto invernadero causados por la actividad humana, gases de efecto invernadero causado por la actividad humana*
- e.g. *calidad exterior del aire, calidad del aire exterior*
- d. Inaccuracies due to style and redundancy
 - e.g. *contaminación del aire al aire libre*
 - e.g. *emisión fugitiva fugaz*
 - e.g. *marea negra por petróleo*
 - e.g. *contaminación por contaminantes orgánicos*
- e. Inaccuracies due to ad hoc translations (modulation)
 - e.g. *desastre ecológico marino, contaminación marina*
 - e.g. *paisajes degradados, contaminación visual*
 - e.g. *exceso de ruido, contaminación acústica*
 - e.g. *envenenamiento industrial, contaminación industrial*

As can be inferred from the classification above there are several structural shifts that also convey a difference in meaning (i.e. inaccuracies, modulations and certain omissions). Cognitive variants occur on the changes mostly affecting nouns, whether in the modifier or in the head, but especially in the latter. As for term opacity, different structures convey more transparent meanings thanks to explicitation. For instance, the preposition *por* is more specific than *de* for making causal relations explicit (e.g. *contaminación por petróleo, contaminación de petróleo*), since *de* is naturally more ambiguous in Spanish.

The most frequent types of variation found in our study were: (1) the omission of articles; (2) the changes in modifiers (reflecting structural or semantic modifications); and (3) the introduction of periphrastic structures, often through participles such as *causado, producido, provocado, inducido, originado, ocasionado*, etc. (e.g. *cambio climático producido/provocado por el hombre*), and relative clauses followed by verbs such as *provocar, causar, contribuir, originar, and producir* (e.g. *emisión de gases que provocan el efecto invernadero*). Inaccuracies, while not that frequent, are worth mentioning,

since they are especially due to the incorrect interpretation of the semantics and internal dependencies of MWTs.

Quite often, several of these types coincide within the same set of variants conveying the same concept, as in *contaminación transfronteriza a gran distancia* (where only structural changes and synonyms apply) (Table 4) or *smog fotoquímico* (where cognitive variants stand out) (Table 5):

contaminación atmosférica transfronteriza a larga distancia (adjective changes)
contaminación atmosférica transfronteriza de larga distancia (preposition changes)
contaminación atmosférica transfronteriza de largo alcance (noun changes)
contaminación atmosférica transfronteriza a largas distancias (number changes)
contaminación atmosférica internacional de largo alcance (adjective changes)
contaminación atmosférica transfronteriza a gran distancia (adjective changes)

Table 4: Spanish variants for *contaminación transfronteriza a gran distancia*.

<i>bruma de verano</i>	<i>niebla fotoquímica</i>
<i>bruma fotoquímica</i>	<i>niebla sucia</i>
<i>bruma industrial de verano</i>	<i>niebla tóxica estival</i>
<i>bruma industrial estival</i>	<i>niebla urbana de verano</i>
<i>contaminación causada por el ozono</i>	<i>nubes de ozono</i>
<i>contaminación de verano</i>	<i>ozono en el aire ambiente</i>
<i>contaminación estival</i>	<i>ozono fotoquímico</i>
<i>contaminación estival por ozono</i>	<i>ozono troposférico</i>
<i>contaminación fotoquímica</i>	<i>polución urbana</i>
<i>contaminación por ozono</i>	<i>smog de Los Ángeles</i>
<i>contaminación veraniega</i>	<i>smog de verano</i>
<i>esmog de ozono</i>	<i>smog estival</i>
<i>esmog de verano</i>	<i>smog fotoquímico</i>
<i>esmog fotoquímico</i>	<i>smog fotoquímico oxidante</i>
<i>neblina estival</i>	<i>smog relacionado con el ozono</i>
<i>neblumo fotoquímico</i>	<i>smog tipo Los Ángeles</i>

Table 5: Spanish variants for *smog fotoquímico*.

Whereas, in Table 4, the variants for *contaminación transfronteriza a gran distancia* usually imply structural changes and synonyms (e.g. *gran* > *larga*; *distancia* > *alcance*; *a larga distancia* > *de larga distancia*; *a larga distancia* > *a largas distancias*, etc.), term variants in Table 5 are quite different. In this

case, cognitive variants are noticeable, which result from the different conceptualizations and modulation of the same concept. For instance, different dimensions are highlighted in the modifiers, whether they are adjectives or nouns. Some of them point to the time when this type of pollution usually occurs (*bruma de verano*, *contaminación de verano*, *contaminación estival*, *esmog de verano* [summer smog]). Others show the city where it was first described (*smog de Los Ángeles*, *smog tipo Los Ángeles* [Los Angeles smog]). Variants can also introduce the agent producing this pollution (*contaminación por ozono*, *esmog de ozono*, *nubes de ozono*, *ozono en el aire ambiente*), or even the process that causes it, the chemical reaction of ozone and light (*bruma fotoquímica*, *contaminación fotoquímica*, *esmog fotoquímico*, *neblumo fotoquímico*). Heads also show cognitive variation, which can also entail changes in the general conceptual category, as in *esmog de ozono* and *ozono fotoquímico* or less complex and less accurate conceptualizations, such as *bruma de verano*, *niebla sucia*, and *polución de verano*. Additional structural aspects are also observed in Table 5, such as the use of the adapted borrowing *esmog* and its non-adapted variant *smog*. SMOG FOTOQUÍMICO is thus a clear example of the richness of term variation.

Therefore, the representation of term variation in terminological resources should be adapted to the different types and consequences of variants. A possible way of covering structural variants (e.g. morphological or morpho-syntactic variants) would be by grouping and ranking them by frequency. However, frequency alone cannot be used as the sole criterion of classification, since other motivations, such as stylistic, cognitive or functional aspects, can be involved in term selection and should also be somehow represented. Consequently, for cognitive variants, frequency as well as the semantic content emphasized by the term should be made explicit. In turn, morphological and morphosyntactic variants should highlight their differences contrastively (as in Table 4). Besides, when necessary, specific markers on term formation devices should also be included (e.g. calque, borrowing, acronym, etc.). However, these issues are not the scope of this study and will be investigated in future research.

5.2. Comparing translation variation across different resources

In line with the findings reported by Sanz Vicente (2011) and Miyata & Kageura (2016), parallel corpora show more translation variants than comparable corpora. Based on our results, 68 out of the 98 concepts were designated with more different terms in parallel corpora, whereas 30 showed more term variants in comparable corpora. In those cases where the comparable corpora provided more variants, there was a slight difference in the number of possible term choices. In contrast, when the parallel corpora showed more variants, the difference amounted to up to 24 additional translations for the same concept. The proliferation of variants in parallel corpora could confirm the hypothesis that translators do not always consider MWTs as a whole unit of understanding but rather as different chunks of independent strings.

In parallel corpora, the use of Spanish variants often depends on the choice made in the source language. For example, *anthropogenic climate change* is usually translated as *cambio climático antropógeno* or *cambio climático antropogénico*, whereas when the same concept is referred to as *human-induced climate change*, most frequent Spanish variants include *cambio climático inducido por el hombre* or *cambio climático provocado por las actividades humanas*. A similar observation can be made regarding synthetic or longer variants in source terms. *Emissions of greenhouse gases* gives rise to more periphrastic choices in Spanish than *greenhouse gas emissions* (e.g. *producción de gases que contribuyen a la creación del efecto invernadero*). However, not only does it occur with regards to synthetic variants, but also in relation to register and transparency. When the source text contains *atmospheric pollution*, it is usually translated as *contaminación atmosférica*, whereas when *air pollution* is chosen, both *contaminación atmosférica* and *contaminación del aire* emerge in Spanish texts. *Contaminación de la atmósfera*, a much less frequent variant in Spanish, only arises when *atmospheric pollution* does in English texts, which would be an example of how parallel corpora do not always show most idiomatic forms of specialized terms. On the contrary, there are also cases where the use of English variants does not give rise to clear-cut cross-linguistic equivalent variants, as in *ozono troposférico*, which is not only used as the literal equivalent of *tropospheric ozone* but also as that of *ground-level ozone* in approximately half the cases where *ground-level ozone* is used in English texts.

Looking at the quantitative results⁴ in more detail, it is worth noting that among the seven resources (Figure 2), OPUS and Google Scholar were those where more translation variants were found (597 and 758 respectively), followed by Linguee (480) (even though not all results are displayed), EurLex (444) and EcoLexicon (450), and finally Termium Plus (105) and IATE (118), which is not surprising since terminological resources usually reflect up to two or three variants.

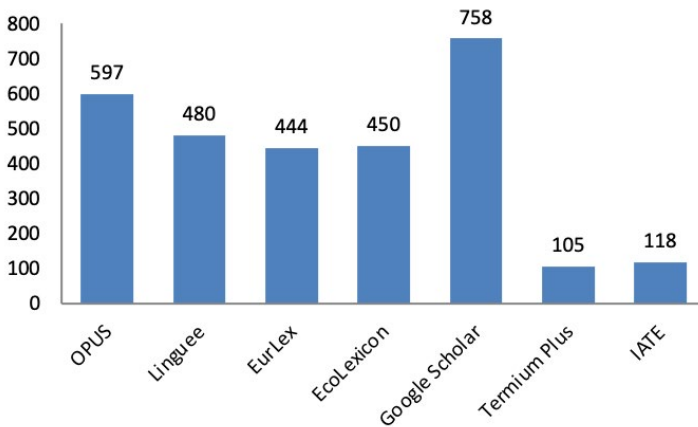


Figure 2: Number of translation variants found in each resource.

Evidently, this does not mean that OPUS and Google Scholar are the best resources to find more reliable results, since reliability would rather depend on the amount of resources containing the same term choices. In this sense, not all translation variants were found in all seven resources; only 25 Spanish terms (pertaining to 25 concepts) were confirmed by all. A staggering 73% of the terms were found only in one of the resources (434 terms) or two of the resources (598 terms) (Figure 3). This means that only 25 concepts have a clearly preferred variant and that most translation variants are not highly lexicalized.

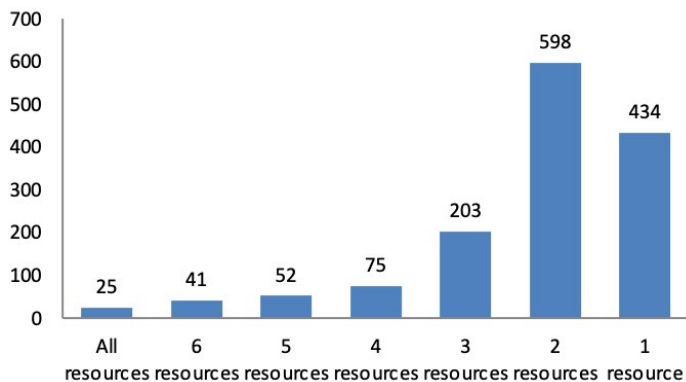


Figure 3: Number of translation variants found in all, several or one of the resources.

The results of analyzing the amount of terms that were found only in one of the resources confirm that parallel corpora contain a higher amount of infrequent or *ad hoc* translation variants (Figure 4).

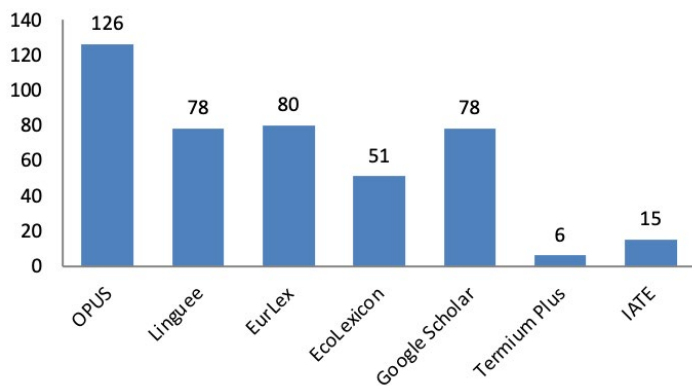


Figure 4: Number of translation variants found only once in each resource.

What is striking is the fact that, although few in number, terminological resources also present very rare translation variants, such as *bióxido de carbono equivalente* (Termium Plus) or *contaminante corpuscular* (IATE). Furthermore, frequency is not always taken into account in these term bases. For example,

Termium Plus only includes the variant *cambio climático mundial* as an equivalent of *global climate change*, although *cambio climático global* exhibits a much higher frequency since the last forty decades, as shown in the Google Ngram Viewer (Figure 5).

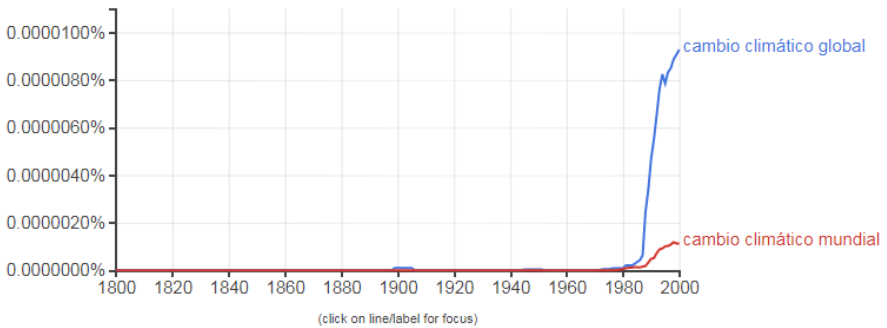


Figure 5: Frequency of *cambio climático global* and *cambio climático mundial* in the Google Ngram Viewer.

6. Conclusions

This study delves into English term variants of MWTs from the environmental domain and their translations into Spanish. The focus was on translation variation and its occurrence in different linguistic resources (i.e. parallel corpora, terminographic resources, and comparable corpora). A typology specifically conceived to characterize translation correspondences of MWTs has been proposed based on omissions, structural shifts, transpositions, expansions and inaccuracies. Additionally, parallel corpora have been found to include more translation variants than other linguistic resources. This proliferation could confirm the hypothesis that translators do not always consider MWTs as a whole unit of understanding but rather as different chunks of independent strings. However, the greater number of translation variants in parallel corpora can also respond to the general texts included in these corpora, even if it was expected that our set of MWTs did not appear very frequently in general texts, given their specialization. Since, to the best of our knowledge, parallel corpora on environmental science are scarce or even non-existent,

an investigation of term variation in this type of corpus should be carried out in the future. Our results highlight the need for an accurate term variation representation in terminographic resources, such as terminological knowledge bases (TKBs). Since they are conceived for different user types and specialized knowledge is not only conveyed in expert-to-expert scenarios, we believe that term variants should be extensively covered in TKBs. However, a balance between thoroughness and over-information must be sought for, because entries presenting 46 different variants (as was found in some of the concepts studied) could be counter-productive. Variation information should thus cover different types of variants and usage-related information, such as frequency and semantic or pragmatic indications. As a future line of research, we plan to implement the findings of this study in the TKB EcoLexicon, where new modules will be offered to compare the use, frequency, and meaning of term variants.

References

- AGUADO DE CEA, Guadalupe & Elena Montiel Ponsoda. (2012) "Term variants in ontologies." In: *XXX Congreso Internacional de AESLA 2012*. Lleida: AESLA, pp. 436-443.
- BENSON, Morton; Evelyn Benson & Robert F. Ilson. (1986) *Lexicographic Description of English*. Studies in Language Companion Series 14. Amsterdam, Philadelphia: John Benjamins.
- BEVILACQUA, Cleci Regina. (2004) *Unidades fraseológicas eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Barcelona: Universitat Pompeu Fabra. PhD thesis.
- BOWKER, Lynne. (1998) "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta* 43:4, pp. 631-651.
- BOWKER, Lynne & Shane Hawkins. (2006) "Variation in the organization of medical terms. Exploring some motivations for term choice." *Terminology* 12:1, pp. 79-110.
- CABEZAS GARCÍA, Melania & Santiago Chambó. (in press) "Multi-word term variation: Prepositional and adjectival complex nominals in Spanish." *RESLA (Revista Española de Lingüística Aplicada)*.
- CABRÉ, María Teresa. (1993) *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida, Empúries.

- CANDEL MORA, Miguel Ángel & María Luisa Carrió Pastor. (2012) "Corpus analysis: a pragmatic perspective on term variation." *RESLA (Revista Española de Lingüística Aplicada)* 2012, pp. 33-50.
- COLLET, Tanja. (2003) "A two-level grammar of the reduction processes of French complex terms in discourse." *Terminology* 9:1, pp. 1-27.
- CORPAS PASTOR, Gloria. (1996) *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, Gloria. (2000) "Fraseología y traducción." In: Salvador, Vicent & Adolf Piquer. (eds.) 2000. *El discurs prefabricat. Estudis de fraseologia teòrica i aplicada*. Estudis Filològics 4. Castelló de la Plana: Servei de Publicacions de la Universitat Jaume I, pp. 49-62.
- DAILLE, Béatrice. (2003) "Conceptual structuring through term variations." Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment. Sapporo, Japan: ACL, pp. 9-16.
- DAILLE, Béatrice. (2017) *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Amsterdam: John Benjamins.
- DAILLE, Béatrice & Emmanuel Morin. (2005) "French-English Terminology Extraction from Comparable Corpora." In: Dale, Robert; Kam-Fai Wong; Jian Su & Oi Yee Kwong. (eds.) 2005. *IJCNLP 2005*. Lecture Notes in Computer Science 3651. Berlin, Heidelberg: Springer, pp. 707-718.
- DÉJEAN, Hervé & Éric Gaussier. (2002) "Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables." *Lexicometrica* 2002, pp. 1-22.
- FABER, Pamela & Pilar León Araúz. (2016) "Specialized knowledge representation and the parameterization of context." *Frontiers in Psychology* 7:196, pp. 1-20.
- FERNÁNDEZ-SILVA, Sabela. (2016) "The Cognitive and Rhetorical Role of Term Variation and its Contribution to Knowledge Construction in Research Articles." *Terminology* 22:1, pp. 52-79.
- FERNÁNDEZ-SILVA, Sabela. (2018) "The cognitive and communicative functions of term variation in research articles: a comparative study in Psychology and Geology." *Applied Linguistics* 2018, pp. 1-23.
- FERNÁNDEZ-SILVA, Sabela; Judit Freixa & María Teresa Cabré. (2009) "The multiple motivation in the denomination of concepts." *Journal of Terminology Science and Research* 20, pp. 1-24.
- FERNÁNDEZ-SILVA, Sabela & Koen Kerremans. (2011) "Terminological variation in source texts and translations: a pilot study." *Meta* 56:2, pp. 318-335.

- FIRTH, John Rupert. (1957) *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- FREIXA, Judit. (2006) "Causes of Denominative Variation in Terminology: A typology proposal." *Terminology* 12:1, pp. 51-77.
- FREIXA, Judit & Sabela Fernández-Silva. (2017) "Terminological variation and the unsaturability of concepts." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 155-181.
- GARCÍA-PAGE, Mario. (2008) *Introducción a la fraseología española*. Barcelona: Anthropos.
- GERZYMISCH-ARBOGAST, Heidrun. (2008) "Fundamentals of LSP Translation." *Mutra* 2, pp. 7-55.
- GIACOMINI, Laura. (2018) "Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary." In: *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*. Ljubljana: EURALEX, pp. 309-318.
- GLÄSER, Rosemarie. (1988) "The grading of idiomaticity as a presupposition for a taxonomy of idioms." In: Hüllen, Werner & Rainer Schulze. (eds.) 1988. *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*. Tübingen: Max Niemeyer, pp. 264-279.
- GLEDHILL, Christopher & Mojca Pecman. (2018) "On Alternating Pre-Modified and Post-Modified Nominals Such As Aspirin Synthesis Versus Synthesis of Aspirin: Rhetorical and Cognitive Packing in English Science Writing." *Fachsprache* 40:1-2, pp. 26-48.
- GREFENSTETTE, Gregory. (2002) "Multilingual corpus-based extraction and the Very Large Lexicon." In: Borin, Lars. (ed.) 2002. *Parallel corpora, parallel worlds*. Amsterdam, New York: Rodopi, pp. 135-149.
- GREGORY, Michael & Susanne Carroll. (1978) *Language and Situation: Language Varieties and their Social Contexts*. London: Routledge.
- HURTADO ALBIR, Amparo. (2001) *Traducción y Traductología: Introducción a la traductología*. Madrid: Cátedra.
- JIMÉNEZ-CRESPO, Miguel Ángel & Maribel Tercedor-Sánchez. (2017) "Lexical variation, register and explicitation in medical translation: A comparable corpus study of medical terminology in US websites translated into Spanish." *Translation and Interpreting Studies* 12:3, pp. 405-426.

- KERREMANS, Koen. (2010) "A comparative study of terminological variation in specialised translation." In: Heine, Carmen & Jan Engberg. (eds.) 2010. *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. Aarhus: Aarhus University, pp. 1-14.
- KERREMANS, Koen. (2016) "Variation in the translation of terms: corpus-driven terminology research." In: Lewandowska-Tomaszczyk, Barbara; Marcel Thelen; Gys-Walt van Egdom; Dirk Verbeeck & Łukasz Bogucki. (eds.) 2016. *Translation and Meaning*. Frankfurt am Main: Peter Lang, pp. 217-227.
- KERREMANS, Koen. (2017) "Towards a resource of semantically and contextually structured term variants and their translations." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 83-108.
- KERREMANS, Koen & Rita Temmerman. (2016) "Finding the (un)expected? Quantitative and qualitative comparisons of term variants and their translations in a parallel corpus of EU texts." In: Corpas Pastor, Gloria & Miriam Seghiri. (eds.) 2016. *Corpus-based Approaches to Translation and Interpreting: from theory to applications*. Frankfurt am Main: Peter Lang, pp. 43-63.
- KILGARRIFF, Adam; Vít Baisa; Jan Bušta; Miloš Jakubíček; Vojtěch Kovář; Jan Michelfeit; Pavel Rychlý & Vít Suchomel. (2014) "The Sketch Engine: ten years on." *Lexicography* 1:1, pp. 7-36.
- LEÓN ARAÚZ, Pilar; Arianne Reimerink & Pamela Faber. (2019) "EcoLexicon and by-products: integrating and reusing terminological resources." *Terminology* 25:2, pp. 222-258.
- LEÓN ARAÚZ, Pilar. (2017) "Term and Concept Variation in Specialized Knowledge Dynamics." In: Drouin, Patrick; Aline Francœur; John Humbley & Aurélie Picton. (eds.) 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam: John Benjamins, pp. 213-258.
- MEL'ČUK, Igor; André Clas & Alain Polguère. (1995) *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- MEYER, Ingrid & Kristen Mackintosh. (1996) "Refining the terminographer's concept-analysis methods: How can phraseology help?" *Terminology* 3:1, pp. 1-26.
- MIYATA, Rei & Kyo Kageura. (2016) "Constructing and Evaluating Controlled Bilingual Terminologies." In: *Proceedings of the Fifth International Workshop on Computational Terminology (Computerm2016)*. Osaka: ACL, pp. 83-93.

- MOGORRÓN-HUERTA, Pedro. (2010) "Analyse du figement et de ses possibles variations dans les constructions verbales espagnoles." *Linguisticae Investigationes* XXXIII:1, pp. 86-152.
- NEWMARK Peter. (1981) *Approaches to Translation*. Oxford: Pergamon.
- NORD, Christiane. (1997) *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome Publishing.
- OLTRA RIPOLL, María Dolores. (2018) "La fraseologia com a tret estilístic en novel·les originals en anglés i les seues traduccions al català." *Caplletra* 65, pp. 95-124.
- PAWLEY, Andrew. (2001) "Phraseology. Linguistics and the Dictionary." *International Journal of Lexicography* 14:2, pp. 122-134.
- PICHT, Heribert. (1991) "Fraseología LSP desde el Punto de Vista Terminológico." *Sendebär* 2, pp. 91-105 (translated from German by Marián Hens Córdoba).
- PICTON, Aurélie. (2011) "Picturing short-period diachronic phenomena in specialised corpora: A textual terminology description of the dynamics of knowledge in space technologies." *Terminology* 17:1, pp. 134-156.
- RAMISCH, Carlos. (2015) *Multiword Expressions Acquisition: A Generic and Open Framework*. Cham: Springer.
- REISS, Katharina & Hans Josef Vermeer. (1984) *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- ROGERS, Margaret. (1997) "Synonymy and equivalence in special-language texts. A case study in German and English texts on Genetic Engineering." In: Trosborg, Anna (ed.) 1997. *Text Typology and Translation*. Amsterdam, Philadelphia: John Benjamins, pp. 217-245.
- SANZ VICENTE, Lara. (2011) *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español*. Salamanca: Universidad de Salamanca. PhD thesis.
- SINCLAIR, John. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- TEMMERMAN, Rita. (2000) *Towards New Ways of Terminology Description: The sociocognitive approach*. Amsterdam, Philadelphia: John Benjamins.
- TIEDEMANN, Jörg. (2012) "Parallel Data, Tools and Interfaces in OPUS." In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: ELRA, pp. 2214-2218.
- VAISA, Vít; Jan Michelfeit; Marek Medved & Miloš Jakubiček. (2016) "European Union Language Resources in Sketch Engine." In: *Proceedings of the Tenth*

- International Conference on Language Resources and Evaluation (LREC'16)*.
Portorož: ELRA, pp. 2799-2803.
- WÜSTER, Eugen. (1968) *The Machine Tool. An Interlingual Dictionary of Basic Concepts*. London: Technical Press.
- ZULUAGA, Alberto. (1975) "La fijación fraseológica." *Thesaurus* XXX:2, pp. 225-248.

BIONOTE / NOTA BIOGRÁFICA

PILAR LEÓN ARAÚZ is a Senior Lecturer at the Faculty of Translation and Interpreting in Granada, where she teaches courses in the fields of Terminology and CAT tools. She received her PhD in Translation and Interpreting from the University of Granada in 2009. She also holds a BA in Applied Foreign Languages from the Université de Provence (France) and a BA in Applied Modern Languages from the University of Northumbria (UK). Her major fields of research include Terminology, Knowledge Representation and Extraction and Corpus Linguistics. She has published papers in international journals such as *Terminology*, *International Journal of Lexicography* and *Frontiers*, and serves on the scientific boards of international journals and conferences.

PILAR LEÓN ARAÚZ es Profesora Titular en la Facultad de Traducción e Interpretación de la Universidad de Granada, donde imparte asignaturas relacionadas con la Terminología y las herramientas de traducción asistida por ordenador. En 2009 obtuvo el Doctorado en Traducción e Interpretación por la Universidad de Granada. Es Licenciada en Lenguas Extranjeras Aplicadas por la Universidad de Provence (Francia) y en Lenguas Modernas Aplicadas por la Universidad de Northumbria (Reino Unido). Sus principales áreas de investigación son la terminología, la representación y extracción del conocimiento, y la lingüística de corpus, sobre las cuales ha publicado artículos en revistas internacionales como *Terminology*, *International Journal of Lexicography* y *Frontiers*. Además, forma parte del comité científico de revistas y congresos internacionales.

MELANIA CABEZAS GARCÍA holds a degree in Translation and Interpreting from the University Pablo Olavide and a Master's degree in Professional Translation from the University of Granada, where she teaches courses in the Bachelor's

Degree in Translation and Interpreting. She received her PhD in Translation and Interpreting from the University of Granada in 2019. She is a member of the LexiCon research group and was awarded a research fellowship by the Spanish Ministry of Education to write her PhD dissertation on the formation, translation, and representation of English and Spanish multiword terms. Her research interests are Terminology, Corpus Linguistics, and Specialized Translation. She has published papers in international journals on Linguistics, Terminology, and Specialized Translation, and serves on the scientific boards of international journals and conferences, such as *LREC* and the *International Journal of Lexicography*.

MELANIA CABEZAS GARCÍA es Graduada en Traducción e Interpretación por la Universidad Pablo de Olavide y Máster en Traducción Profesional por la Universidad de Granada, donde imparte docencia en el Grado en Traducción e Interpretación. Obtuvo el Doctorado en Traducción e Interpretación por la Universidad de Granada en 2019. Perteneció al grupo de investigación LexiCon y contó con una beca de investigación predoctoral otorgada por el Ministerio de Educación de España para realizar su tesis doctoral, en la que estudió la formación, traducción y representación de los términos compuestos en inglés y español. Sus áreas de investigación son la terminología, la lingüística de corpus y la traducción especializada. Ha publicado artículos en revistas internacionales sobre lingüística, terminología y traducción especializada, y forma parte del comité científico de revistas y congresos internacionales, como *LREC* o *International Journal of Lexicography*.